Generative Anchor-Guided Federated Domain Generalization for Heterogeneous Pan Cancer Image Analysis

Qiangguo Jin[1], Yufei Shen[1], Linkuan Zhou[1], Fei Guo[1,7], Junlin Xu[2], Ping Xuan[3], Ran Su[4], Leyi Wei[5], and Cong Cong[6*]

[1] School of Software, Northwestern Polytechnical University, Shaanxi, China
ggking@nwpu.edu.cn (Q. Jin)
[2] School of Computer Science and Technology, Wuhan University of Science and Technology, Hubei, China
[3] School of Cyberspace Security, Hainan University, Hainan, China
[4] School of Computer Software, Tianjin University, Tianjin, China
[5] Centre for Artificial Intelligence driven Drug Discovery, Faculty of Applied Science, Macao Polytechnic University, Macao Special Administrative Region of China
[6] School of Computing, Macquarie University, New South Wales, Australia
[7] School of Information Technology, Suzhou Institute of Trade and Commerce, Jiangsu, China

Abstract. Federated learning (FL) has emerged as a key paradigm for collaborative cancer subtyping classification, which enables the development of diagnostic models across distributed institutions while preserving data privacy. Despite the emergence of federated approaches, their performance is often constrained by the assumption of a unified label space, which fails to account for the heterogeneous diagnostic criteria encountered in real-world multi-center collaborations. To address this, we propose FedGMA, a framework designed for federated domain generalization under disjoint semantics. Our method introduces generative anchor-based semantic decoupling (GASD), which projects latent prompts into task-specific geometric coordinates to decouple feature learning from local label definitions. We further propose geometric consistency regularization (GCR) to regularize the latent space using a memory bank, ensuring representational compactness and separability under limited batch sizes. To evaluate our approach under heterogeneous label spaces, We construct a large-scale benchmark with 4,796 whole slide images across 8 cancer types. FedGMA achieves 76.42% average AUC, outperforming the state-of-the-art by 4.96% on unseen target domains. Code is available at https://github.com/BioMedIA-repo/fedGMA_pan_cancer.

Keywords: Federated Domain Generalization · Heterogeneous Label Spaces · Cross-cancer Subtype.

1 Introduction

Robust computational pathology algorithms rely on diverse, large-scale datasets for clinical-grade performance, yet privacy regulations prohibit centralizing sen-

sitive patient records. Federated learning (FL) resolves this dilemma by enabling collaborative training without data sharing, allowing models to learn from global distributions while keeping data local. Previous studies [8,13,14,16,26] have successfully applied FL to pathology datasets for segmentation and classification, effectively handling scanner-induced domain shifts through aggregated model updates. These approaches demonstrate that collaborative learning narrows domain gaps from image acquisition differences, establishing FL as the prevailing paradigm for multi-center pathology analysis.Even recent advances such as FedDFP [2] and FedHD [9], which improve communication efficiency and model heterogeneity tolerance via deep feature prompts and feature distillation respectively, inherit this assumption of a shared label space.

However, general FL methods [10,18,21] predominantly assume a unified label space across all clients. This premise fails to account for real-world pan-cancer collaborations, where specialized centers utilize heterogeneous label spaces. These spaces consist of disjoint category sets that are specific to local diagnostic tasks and distinct from multi-label classification. Although multi-task and personalized FL approaches [12,20,23] can accommodate such disjoint tasks, they prioritize optimizing client-specific performance for personalization rather than constructing a universal representation for unseen target domains. Such a localized objective misaligns with the pan-cancer ambition of developing diagnostic tools capable of adapting to new or rare cancer types. From a technical perspective, these methods rely on static classification heads tightly coupled with the shared feature extractor. Aggregating gradients from these semantically conflicting heads induces interference in the representation learning process by forcing the encoder to resolve contradictory optimization directions. Consequently, such aggregation precipitates negative transfer [25], where the global model overfits to local label semantics instead of capturing generalized pathological features essential for unseen domains. Prototype-based and anchor-based decoupling methods such as FedSA [27] and FedLSA [3] follow a similar spirit. They replace averaged prototypes with learnable semantic anchors to mitigate classifier conflicts, yet these anchors remain class-indexed. When each client diagnoses an entirely distinct set of cancer subtypes, no shared class-level correspondence exists to serve as an anchoring target.

To bridge this gap, we draw the insight from recent pathology foundation models [1, 22] revealing that distinct cancer types share fundamental histomorphological primitives invariant across organs [4]. Based on this, we hypothesize that generalizable learning hinges on disentangling feature geometry, defined as the intrinsic topology of histological patterns, from site-specific semantics derived from disjoint local labels. By shifting the optimization objective from fitting local labels to constructing a coherent feature geometry, a global model can capture universal pathological representations without requiring a unified label space. We instantiate this principle through a shared hypernetwork that synthesizes generative anchors from semantic-free latent codes, guiding feature alignment toward morphological rather than semantic targets.

Driven by this insight, we propose federated geometric metric adaptation (FedGMA), a framework that unifies feature geometry across disparate cancer types. By leveraging pan-cancer histomorphological invariants, FedGMA facilitates robust adaptation to unseen target domains under the challenging setting of FedDG with heterogeneous label spaces. To decouple representation learning from conflicting local labels, we introduce generative anchor-based semantic decoupling (GASD). In this module, a shared hypernetwork transforms semantic-agnostic latent prompts into client-specific geometric anchors. Since these anchors derive from noise rather than clinical metadata, they serve as pure geometric reference points. Aligning features with these semantic-free references effectively circumvents site-specific label conflicts, reframing classification as a geometric registration problem. Furthermore, to prevent feature collapse under limited batch settings, we propose geometric consistency regularization (GCR). GCR utilizes a memory bank to retrieve historical features as negative samples, explicitly prompting intra-class compactness and inter-class separability, thereby preserving the discriminative topology of the learned feature space. Our main contributions are summarized as follows:

- We introduce generative anchor-based semantic decoupling (GASD), a paradigm where a shared hypernetwork synthesizes consistent geometric anchors from latent prompts, enabling discriminative learning without encoding conflicting local label semantics.
- We propose geometric consistency regularization (GCR) to stabilize optimization under limited batch constraints. By retrieving historical negatives via a memory bank, GCR actively enforces feature compactness and separability within the generated metric space.
- We construct a large-scale pan-cancer benchmark spanning 8 disjoint subtypes across 4,796 WSIs. Extensive experiments demonstrate that FedGMA significantly outperforms state-of-the-art baselines on unseen target domains.

2 Dataset Collection and Benchmark Construction

To systematically evaluate federated domain generalization in highly heterogeneous environments, we construct a standardized benchmark using multi-source WSIs. Following the cancer subtype selection strategy of CPath-Omni [19], we focus on the same eight major cancer types (RCC, NSCLC, BRCA, UCEC, THCA, ESCA, BLCA, TGCT). Fig.2 illustrates the histological subtype distribution within each cancer site. We preprocess all WSIs using the CLAM pipeline [15] to standardize the input for our model, employing HSV thresholding for background removal and 10× patch extraction.

To construct the federated environment, we assign each cancer subtype to an independent client to simulate extreme data heterogeneity inherent in real-world clinical scenarios. Formally, we consider a system with $K = 8$ sites, where the k-th site holds a local dataset $\mathcal{D}_k = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_k}$ of WSIs with labels $y_i \in \mathcal{Y}_k$. A critical characteristic of this benchmark is the complete mutual exclusivity of

label spaces, such that for any two distinct sites i and j, the intersection of their label sets is empty, $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$. This configuration challenges the conventional assumption of shared label spaces in federated learning and provides a rigorous testing ground for validating model generalization on unseen target domains.

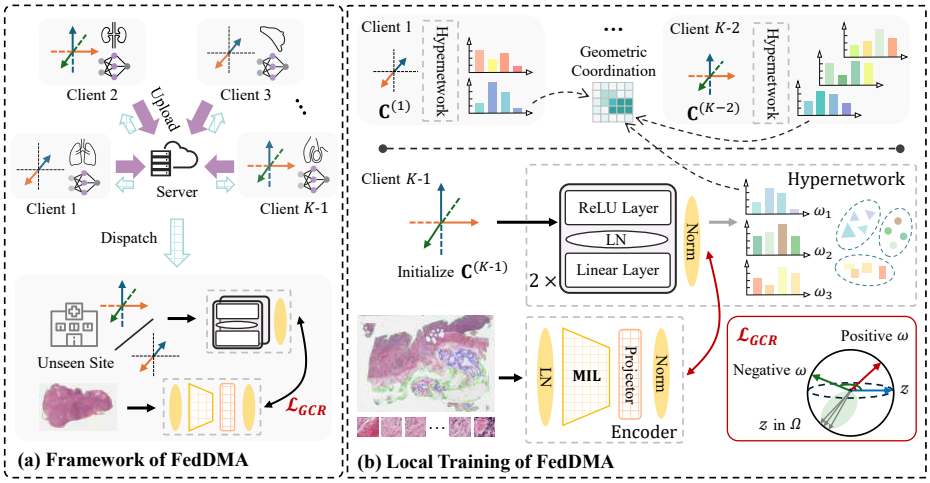


Fig. 1: The framework of FedGMA and the architecture of the client model.

3 Methodology

3.1 Overall FedGMA Framework

The proposed FedGMA framework is illustrated in Fig.1. The overall federated pipeline (Fig.1a) coordinates collaborative learning across heterogeneous clients to generalize to unseen target domains. Specifically, the local training process (Fig.1b) optimizes the shared encoder and hypernetwork via generative anchor-based semantic decoupling (GASD) and geometric consistency regularization (GCR), learning a pan-cancer geometric space. Subsequently, the server aggregates the updated encoder and hypernetwork parameters via weighted averaging proportional to local dataset sizes. The aggregated global model is then broadcast back to the clients for the next round of local updates, completing the federated training loop.

During the generalization phase, the final global model is transferred to unseen target domains for fine-tuning, where the learned geometric priors enable rapid adaptation with minimal semantic drift. To ensure framework consistency, the fine-tuning objective strictly aligns with the federated training loss.

3.2 Generative Anchor-Based Semantic Decoupling

To circumvent the systemic constraint where label spaces are disjoint across medical institutions, we establish a decoupling paradigm that separates feature extraction from specific label semantics. Instead of optimizing a fixed classification matrix, the framework learns a universal mapping function that projects semantic-free priors into discriminative feature representations. For client k with N_k WSIs, we first initialize a matrix of static latent prompts $\mathbf{C}^{(k)}$. This matrix is constructed to satisfy strict orthogonality constraints, ensuring maximal separability in the initial latent space, formally defined as:

$$\mathbf{C}^{(k)} = [c_1, c_2, \ldots, c_{N_k}]^\top \in \mathbb{R}^{N_k \times d_{in}}, \quad \text{s.t.} \quad \mathbf{C}^{(k)}(\mathbf{C}^{(k)})^\top = \mathbf{I}_{N_k}, \tag{1}$$

where $\mathbf{I}_{N_k}$ denotes the identity matrix of size N_k, d_{in} denotes predefined input dimension of latent prompts. All orthogonal codes remain frozen during training.

To bridge these latent priors with visual semantics, the framework employs a globally shared hypernetwork Φ to transform the prompts $\mathbf{C}^{(k)}$ into dynamic anchors, which serve as geometric targets for the image features z extracted by the encoder $F(\cdot;\theta_f)$. To mitigate feature instability from scanner artifacts, we constrain the interaction space to a unit hypersphere by performing L_2 normalization on both z and generated anchors. A globally shared hypernet $\Phi(\cdot;\theta_h)$ transforms individual rows c_i from $\mathbf{C}^{(k)}$ into high-dimensional dynamic anchors:

$$\omega_i = \Phi(c_i;\theta_h) = \sigma_2\left(\mathrm{LN}\left(W_2 \cdot \sigma_1\left(\mathrm{LN}\left(W_1 \cdot c_i + b_1\right)\right) + b_2\right)\right), \tag{2}$$

where $\omega_i \in \mathbb{R}^D$ represents the generative discriminative anchor for the i-th class, and $\{W_1, W_2, b_1, b_2\} \subset \theta_h$ are the learnable meta-parameters. The predictive distribution follows a temperature-scaled softmax function:

$$p(y|x) = \frac{\exp((s(z,\omega_y)/\tau))}{\sum_{j=1}^{N_k}\exp((s(z,\omega_j)/\tau)}, \tag{3}$$

where $s(\cdot,\cdot)$ denotes cosine similarity. This hyperspherical projection ensures hypernet learns a consistent geometric dictionary, mapping morphologically similar categories into proximal regions irrespective of institutional labels. During the generalization phase, a new orthogonal prompt $\mathbf{C}^{(target)}$ is initialized for unseen domain categories, enabling the pretrained Φ to synthesize task-specific anchors for rapid fine-tuning with minimal semantic alignment.

3.3 Geometric Consistency Regularization

While the generative anchors established in GASD provide class-specific geometric centers, relying solely on point-to-point alignment is insufficient to shape a robust manifold. To consolidate feature-anchor alignment and stabilize optimization under batch size limitations, We introduce a geometric consistency regularization. A first-in-first-out memory bank $\mathcal{M}$ serves as a dense proxy for

the global distribution, contrasting the current sample against historical negatives. The loss function is formulated as:

$$\mathcal{L}_{\mathrm{GCR}} = -\log \frac{\exp(s(z, \omega^+)/\tau)}{\sum_{r \in \{\omega^+\} \cup \Omega_{\mathrm{neg}} \cup \mathcal{M}} \exp(s(z, r)/\tau)}, \tag{4}$$

where ω^+ is the positive anchor. The negative set $\mathcal{R}$ integrates intra-task negative anchors $\Omega_{\mathrm{neg}} = \{\omega_j \mid j \neq y\}$ and historical embeddings from $\mathcal{M}$.

By maximizing the margin between the aligned feature z and this comprehensive set of geometric pivots, the objective compels the encoder to learn a compact and low-entropy representation. Crucially, by enforcing geometric uniformity locally, this objective prevents client encoders from overfitting to highly disparate local label definitions. Such structural alignment ensures that the server aggregates feature extractors with a compatible geometric foundation, thereby yielding a robust global model poised for explicitly reliable and effective fine-tuning on unseen target domains.

4 Experiments

4.1 Implementation Details

We implement FedGMA using PyTorch and conduct training on a single NVIDIA RTX 4090D GPU. For evaluation, we extract patch-level embeddings using an ImageNet-pretrained ResNet50 [5, 17] and aggregate them via an ABMIL [7] backbone. Each center partitions its data into training, validation, and testing sets at a 7:1:2 ratio. Under the leave-one-domain-out protocol, source clients utilize the 70% and 10% splits for local optimization and early stopping respectively, whereas the unseen target domain employs the 10% split as a support set for fine-tuning and the remaining 90% as a query set for evaluation. The temperature coefficient τ is fixed at 0.07 to calibrate the discriminative sharpness and the memory bank $\mathcal{M}$ holds 32 samples.

Federated training is performed for 20 communication rounds, with each client executing 20 local epochs per round. We utilize the SGD optimizer with a learning rate of $1e-4$ and a weight decay of $5e-3$. Subsequent fine-tuning rapidly adapts the global model for 10 epochs. All experiments are conducted under fixed random seeds and settings for fair comparison.

4.2 Performance Analysis and Component Evaluation

Comparison with State-of-the-Arts Methods. We compare FedGMA with several state-of-the-art methods, including general federated learning paradigms (FedAvg [16], FedProx [13], MOON [11]), domain generalization methods (FedAvg w/RSC [6], FedDA [24]), and personalized approaches (FedProto [20], FedAdapterProto [20]), alongside an Only-finetune baseline trained solely on target data. As shown in Table 1, each column represents the result when the

corresponding cancer type acts as the unseen target domain. We report AUC as the primary metric.

Standard FL methods struggle with label heterogeneity, performing comparably to Only-finetune at approximately 69.47% average AUC. While DG methods like FedAvg w/RSC demonstrate slight improvements, their performance is hampered by semantic misalignment. Similarly, prototype-based approaches such as FedAdapterProto raise the average AUC to approximately 71.5%, but remain constrained by the lack of explicit geometric regularization. In contrast, our method achieves a superior 76.42% average AUC, consistently outperforming the second-best method across all eight target configurations.

Table 1: Results on the benchmarks under the FedDG setting. The best results are highlighted in bold, and the second-best results are underlined. † indicates statistically significant difference compared to FedGMA ($p < 0.01$, Wilcoxon signed-rank test).

Method	BRCA	NSCLC	TGCT	BLCA	RCC	UCEC	THCA	ESCA	**Avg.**
Only-finetune†	69.87	75.03	77.76	52.71	84.19	55.08	62.73	78.40	69.47
FedAvg†	75.98	74.33	77.14	51.83	82.35	54.78	64.19	76.71	69.66
FedProx†	75.97	74.53	76.85	51.72	81.72	54.66	67.60	77.46	70.06
MOON†	75.98	74.33	76.47	51.83	82.77	54.78	64.19	76.71	69.63
FedAvg w/RSC†	76.21	73.56	76.31	50.99	83.33	55.00	65.00	77.37	69.72
FedDA†	74.76	74.09	76.72	51.02	82.87	55.95	64.24	75.90	69.44
FedProto†	75.97	73.58	76.58	51.28	81.91	55.01	62.66	77.61	69.32
FedAdapterProto†	71.45	<u>77.63</u>	78.05	51.90	<u>87.67</u>	56.14	<u>70.74</u>	78.09	71.46
*FedGMA**	<u>76.59</u>	76.77	<u>78.25</u>	<u>53.26</u>	86.05	<u>58.04</u>	67.08	<u>79.97</u>	<u>72.00</u>
FedGMA	**77.67**	**77.78**	**79.58**	**59.74**	**91.42**	**64.61**	**71.52**	**89.02**	**76.42**

*FedGMA** means FedGMA w/o GCR

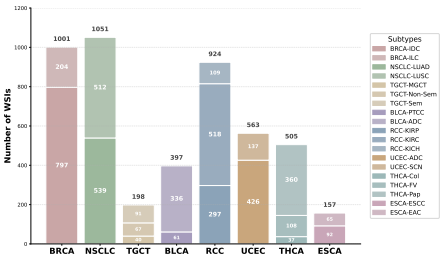
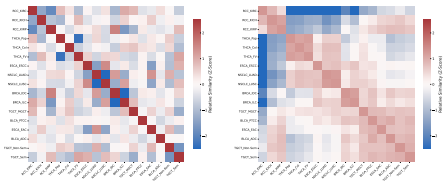


Fig. 2: Data composition across eight cancer types.



(a) FedAvg (b) Ours

Fig. 3: Heatmap comparison of anchor similarity across domains.

Ablation Study. To evaluate the individual contribution of the key modules, we conducted an incremental analysis as shown at the bottom of Table 1. When the framework utilizes only GASD, denoted as ours w/o GCR, it already achieves an average AUC of 72.00%, outperforming existing federated paradigms. This result confirms our hypothesis that generating dynamic anchors from latent prompts effectively circumvents the conflicts inherent in coupled classification heads. The full version of our framework further integrates GCR, boosting average AUC to 76.42%. In summary, the synergy between GASD and GCR enables effective pan-cancer knowledge leverage, establishing a new benchmark for federated domain generalization in pathology.

Interpretability of Cross-Domain Semantic Coordination. To examine the internal classification logic, we visualize the similarity heatmap of generated anchors using z-score normalization. As shown in Fig.3, using UCEC as the target domain to visualize client anchors, our framework exhibits structured geometric coordination absent in the fragmented space of the FedAvg baseline. A significant finding is the convergence of adenocarcinoma subtypes across organ sites, where BLCA-ADC and ESCA-ADC are mapped into highly proximal latent regions despite disjoint label sets. This coordination suggests that the hypernet has captured invariant histomorphological primitives such as glandular structures. Furthermore, the geometric isolation of RCC (KIRC, KIRP, KICH) demonstrates that the decoupling mechanism maintains high specificity for unique pathological patterns. These results confirm that our framework unifies the classification space by grounding it in universal morphological descriptors.

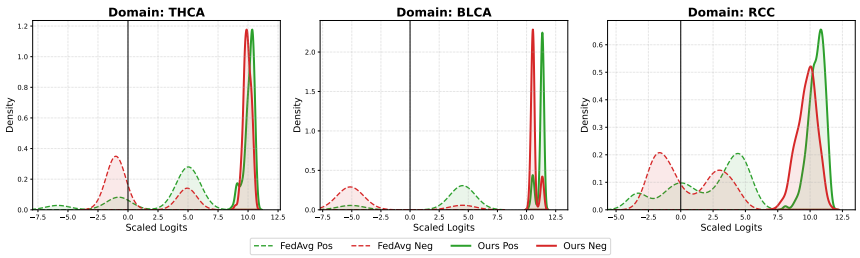


Fig. 4: KDE Comparison of Geometric Feature Distribution on Unseen Domains

Analysis of Geometric Feature Distribution. We further evaluate representational precision on unseen test domains to verify the capture of invariant pathological primitives. A generalized model is expected to yield features highly aligned with generated anchors which manifests as sharp and high-confidence distributions. Evidence from Fig.4 and Table 2 indicates that FedGMA consistently shifts positive distributions toward the high-logit regime. Consequently, the mean alignment magnitude $M_{pos} = \mathbb{E}[s(z_i, \omega^+)/\tau]$ reaches significantly higher levels

Table 2: Quantitative characterization of the feature distributions visualized in Fig.4 across eight unseen target domains. ↓ indicates lower is better, ↑ indicates higher is better.

		BRCA	NSCLC	TGCT	BLCA	RCC	UCEC	THCA	ESCA
FedAvg	σ^2_{pos}	8.82	1.44	2.47	12.28	8.04	5.90	11.67	3.12
	M_{pos}	2.16	0.17	0.49	3.00	2.03	1.63	3.00	0.35
Ours	σ^2_{pos} ↓	**0.01** ↓	**0.16** ↓	**0.09** ↓	**0.10** ↓	**0.38** ↓	**0.01** ↓	**0.18** ↓	**0.22** ↓
	M_{pos} ↑	**12.15** ↑	**11.09** ↑	**4.81** ↑	**11.21** ↑	**10.50** ↑	**6.64** ↑	**10.11** ↑	**8.25** ↑

compared to the baseline, leveraging the features z and anchors ω defined in Sec.3.2. Simultaneously, the intra-class variance $\sigma^2_{pos} = \mathbb{E}[(l_i - M_{pos})^2]$ for instance logits l_i is drastically minimized, reaching 0.01 in the BRCA domain. These high-alignment and low-variance statistics confirm that FedGMA effectively suppresses representational noise to establish compact feature boundaries, providing a solid geometric foundation for superior adaptation performance.

5 Conclusion

We propose FedGMA to address the challenges of federated domain generalization under heterogeneous label spaces in computational pathology. We first construct a standardized multi-cancer benchmark comprising 4,796 WSIs across eight disjoint label spaces. By introducing the GASD and GCR modules, we successfully decouple the learning of invariant histomorphological primitives from disparate local clinical semantics. Our framework leverages pan-cancer knowledge to learn a universal feature encoder that generalizes across diverse organ sites. Experimental results demonstrate that FedGMA significantly outperforms existing methods.

Acknowledgments. This work was supported by the National Natural Science Foundation of China [Grant No. 62572401 and No. 62322112], the Key Research and Development Program of Shaanxi [Program No. 2025SF-YBXM-424], and the Fundamental Research Funds for the Central Universities.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (2024)

2. Cong, C., Song, Y., Di Ieva, A., Chou, A., J. Gill, A., Coiera, E., Liu, S.: Flexible and scalable federated learning with deep feature prompts for digital pathology. npj Digital Medicine (2026)
3. Fu, L., Huang, S., Lai, Y., Liao, T., Zhang, C., Chen, C.: Beyond federated prototype learning: Learnable semantic anchors with hyperspherical contrast for domain-skewed data. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 16648–16656 (2025)
4. Fu, Y., Jung, A.W., Torne, R.V., Gonzalez, S., Vöhringer, H., Shmatko, A., Yates, L.R., Jimenez-Linan, M., Moore, L., Gerstung, M.: Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nature Cancer **1**(8), 800–810 (2020)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
6. Huang, Z., Wang, H., Xing, E.P., Huang, D.: Self-challenging improves cross-domain generalization. In: European Conference on Computer Vision. pp. 124–140. Springer (2020)
7. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning. pp. 2127–2136. PMLR (2018)
8. Jin, H., Liu, S., Cong, C., Feng, Q., Liu, Y., Huang, L., Hu, Y.: Fedwsidd: Federated whole slide image classification via dataset distillation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 178–188. Springer (2025)
9. Jing, L., Cong, C., Chen, Y., Cao, Y.: Federated distillation for whole slide image via gaussian-mixture feature alignment and curriculum integration. arXiv preprint arXiv:2605.00578 (2026)
10. Li, B., Liu, Z., Shao, L., Qiu, B., Bu, H., Tian, J.: Point transformer with federated learning for predicting breast cancer her2 status from hematoxylin and eosin-stained whole slide images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 3000–3008 (2024)
11. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10713–10722 (2021)
12. Li, T., Hu, S., Beirami, A., Smith, V.: Ditto: Fair and robust federated learning through personalization. In: International Conference on Machine Learning. pp. 6357–6368. PMLR (2021)
13. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems **2**, 429–450 (2020)
14. Lu, M.Y., Chen, R.J., Kong, D., Lipkova, J., Singh, R., Williamson, D.F., Chen, T.Y., Mahmood, F.: Federated learning for computational pathology on gigapixel whole slide images. Medical Image Analysis **76**, 102298 (2022)
15. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021)
16. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics. pp. 1273–1282. PMLR (2017)
17. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-

performance deep learning library. Advances in Neural Information Processing Systems **32** (2019)
18. Raswa, F.H., Lu, C.S., Wang, J.C.: Histofs: Non-iid histopathologic whole slide image classification via federated style transfer with roi-preserving. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30251–30260 (2025)
19. Sun, Y., Si, Y., Zhu, C., Gong, X., Zhang, K., Chen, P., Zhang, Y., Shui, Z., Lin, T., Yang, L.: Cpath-omni: A unified multimodal foundation model for patch and whole slide image analysis in computational pathology. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 10360–10371 (2025)
20. Tan, Y., Long, G., Liu, L., Zhou, T., Lu, Q., Jiang, J., Zhang, C.: Fedproto: Federated prototype learning across heterogeneous clients. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 8432–8440 (2022)
21. Ogier du Terrail, J., Leopold, A., Joly, C., Béguier, C., Andreux, M., Maussion, C., Schmauch, B., Tramel, E.W., Bendjebbar, E., Zaslavskiy, M., et al.: Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. Nature Medicine **29**(1), 135–146 (2023)
22. Xiang, J., Wang, X., Zhang, X., Xi, Y., Eweje, F., Chen, Y., Li, Y., Bergstrom, C., Gopaulchan, M., Kim, T., et al.: A vision–language foundation model for precision oncology. Nature **638**(8051), 769–778 (2025)
23. Yang, J., Ma, Y., Yu, J.G., Gao, F., Yang, S., Cai, D., Wang, J., Wang, L.: Bridging knowledge discrepancy in retinal image analysis through federated multi-task learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 35–44. Springer (2025)
24. Zhang, R., Xu, Q., Yao, J., Zhang, Y., Tian, Q., Wang, Y.: Federated domain generalization with generalization adjustment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3954–3963 (2023)
25. Zhang, W., Deng, L., Zhang, L., Wu, D.: A survey on negative transfer. IEEE/CAA Journal of Automatica Sinica **10**(2), 305–329 (2022)
26. Zhang, Y., Chen, F., Qi, Y., Yang, G., Fu, H.: PathFL: Multi-Alignment Federated Learning for Pathology Image Segmentation. arXiv preprint arXiv:2505.22522 (2025)
27. Zhou, Y., Qu, X., You, C., Zhou, J., Tang, J., Zheng, X., Cai, C., Wu, Y.: Fedsa: A unified representation learning via semantic anchors for prototype-based federated learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 23009–23017 (2025)